

The statistical challenge of constraining the low-mass IMF in Local Group dwarf galaxies

Kareem El-Badry,¹[★] Daniel R. Weisz,¹ and Eliot Quataert¹

¹*Department of Astronomy, University of California, Berkeley, CA, USA.*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We use Monte Carlo simulations to explore the statistical challenges of constraining the characteristic mass (m_c) and width (σ) of a lognormal sub-solar initial mass function (IMF) in Local Group dwarf galaxies using direct star counts. For a typical Milky Way (MW) satellite ($M_V = -8$), jointly constraining m_c and σ to a precision of $\lesssim 20\%$ requires that observations be complete to $\lesssim 0.2M_\odot$, if the IMF is similar to the MW IMF. A similar statistical precision can be obtained if observations are only complete down to $0.4M_\odot$, but this requires measurement of nearly $100\times$ more stars, and thus, a significantly more massive satellite ($M_V \sim -12$). In the absence of sufficiently deep data to constrain the low-mass turnover, it is common practice to fit a single-sloped power law to the low-mass IMF, or to fit m_c for a lognormal while holding σ fixed. We show that the former approximation leads to best-fit power law slopes that vary with the mass range observed and can largely explain existing claims of low-mass IMF variations in MW satellites, even if satellite galaxies have the same IMF as the MW. In addition, fixing σ during fitting leads to substantially underestimated uncertainties in the recovered value of m_c (by a factor of ~ 4 for typical observations). If the IMFs of nearby dwarf galaxies are lognormal and *do* vary, observations must reach down to $\sim m_c$ in order to robustly detect these variations. The high-sensitivity, near-infrared capabilities of JWST and WFIRST have the potential to dramatically improve constraints on the low-mass IMF. We present an efficient observational strategy for using these facilities to measure the IMFs of Local Group dwarf galaxies.

Key words: stars: mass function – stars: low-mass – methods: statistical

1 INTRODUCTION

A persistent uncertainty in extragalactic astronomy is whether the stellar initial mass function (IMF) varies with environment and/or cosmic time. Many theoretical models of star formation hold that the shape of the IMF is set by local variables such as the Mach number in molecular clouds (e.g. Padoan et al. 1997; Hopkins 2013), the local Jeans length (e.g. Larson 1998; Narayanan & Davé 2012), or the gas pressure in star-forming regions (e.g. Krumholz 2011; Krumholz et al. 2016), and thus predict that the IMF for a particular star formation event should depend to some extent on the local gas density and metallicity, and on the instantaneous star formation rate (see Offner et al. 2014, for a review).

On the other hand, most observational studies have measured mass functions consistent with a universal IMF across a wide variety of environments (see Bastian et al.

2010, for a review). Some works *have* found integrated light signatures of low-mass IMF variations in the centers of massive early-type galaxies (van Dokkum 2008; Conroy & van Dokkum 2012; Cappellari et al. 2012; Smith 2014; van Dokkum et al. 2016), but these results require further interpretation (McConnell et al. 2016; Coulter et al. 2017). Most IMF studies carried out with direct star counts have found little evidence of IMF variation in the Local Group (Bessell & Stringfellow 1993; Wyse et al. 2002; Chabrier 2003a; Covey et al. 2008; Bastian et al. 2010; Bochanski et al. 2010).

Dwarf spheroidal and ultra-faint galaxies in the Local Group provide a natural laboratory in which to search for environmental variations in the IMF. Compared to the Milky Way (MW), Local Group satellites have extremely low stellar densities, low metallicities ($[\text{Fe}/\text{H}] \sim -2$), high alpha abundances ($[\alpha/\text{Fe}] \sim 0.4$), and high mass-to-light ratios ($M/L \sim 100$, in solar units) (Tolstoy et al. 2009; Frebel & Norris 2015). In addition, many (though not all) Local Group dwarf galaxies exhibit relatively simple star forma-

[★] E-mail: kelbadry@berkeley.edu

tion histories, having formed all their stars within $\lesssim 2$ Gyr, before $z \sim 3$ (Brown et al. 2012; Vargas et al. 2013; Webster et al. 2015; Weisz et al. 2014). One might thus expect systematic IMF variation between Local Group dwarf galaxies and the MW.

Observational searches for IMF variations in MW satellites have thus far yielded mixed results: Grillmair et al. (1998) and Wyse et al. (2002) found IMFs in the Draco and Ursa Minor dwarf spheroidal galaxies that were roughly consistent with the IMF measured in the MW and in its globular clusters, and Kalirai et al. (2013) found the IMF of the Small Magellanic Cloud (SMC) to be only marginally shallower than that of the MW. On the other hand, Geha et al. (2013) found the IMFs of the Hercules and Leo IV ultra-faint dwarf galaxies to be significantly shallower (more bottom-light) than the canonical MW IMF.

All published attempts to measure the IMFs of nearby dwarf galaxies have been limited at low masses by observational sensitivity. At a distance of $D = 150$ kpc, Hubble Space Telescope (HST) photometric studies become incomplete below stellar masses of $\sim 0.5 M_\odot$. This is significantly more massive than the characteristic mass at which the canonical MW IMF turns over and begins to significantly deviate from a power law ($m_c \sim 0.22 M_\odot$). No published observational studies have detected a turnover in the mass functions of nearby dwarf galaxies, and they have thus all opted to fit a simple power law IMF. However, measuring the turnover in the IMF (if indeed it exists) is particularly desirable, as it would provide a direct test of theoretically predicted scalings of m_c with metallicity and density.

The statistical power of resolved stellar population studies to constrain the low-mass IMF will improve significantly with the introduction of the James Webb Space Telescope (JWST), whose larger collecting area and improved near-infrared sensitivity will allow observations to probe the IMFs of Local Group galaxies to lower masses than past HST observations. In the nearest MW satellites, JWST will efficiently observe down to the hydrogen burning limit, making it possible to constrain the characteristic mass and slope of the IMF with unprecedented precision.

In this paper, we use Monte Carlo simulations to investigate how the accuracy to which the IMF can be recovered from star counts scales with the size and limiting magnitude of the observed stellar sample. We demonstrate that HST resolved stellar population studies do not have the statistical power to robustly detect IMF variations in most Local Group dwarf galaxies if variation occurs primarily at lower masses. However, we find that similar studies with JWST will have the potential to definitively quantify (or rule out) low-mass IMF variations for galaxies within ~ 100 kpc.

We organize this paper as follows. In Section 2, we describe how we fit the IMF from a sample of masses or a color-magnitude diagram (CMD). In Section 3, we explore how accurately the parameters of a MW-like lognormal IMF can be constrained with observations reaching a range of limiting magnitudes. We then investigate the systematics introduced by simplifications which are commonly made in the absence of sufficiently deep observations, such as fitting a single power law or holding one of the parameters of a lognormal fixed during fitting. In Section 4, we investigate what observations are required to detect variation in the IMF. In Section 5, we discuss how future observations with

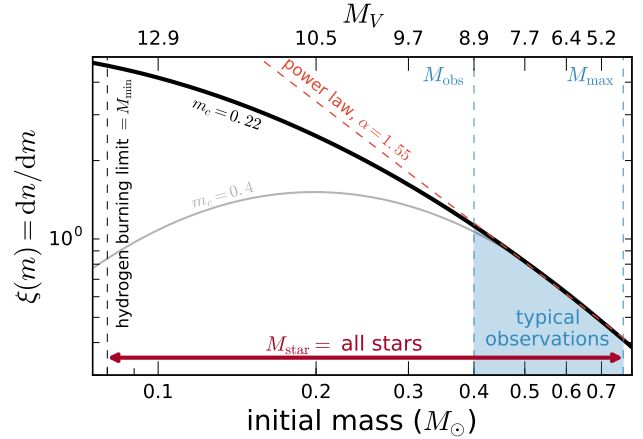


Figure 1. Schematic illustration of the challenge of constraining the low-mass IMF. Observations sample the IMF between M_{obs} , the lowest stellar mass which can be reached by observations, and M_{max} , the highest stellar mass which remains on the main sequence. The total mass M_{star} corresponds to all stars between M_{min} and M_{max} , many of which cannot be observed. When observations do not reach the low-mass turnover, both power law IMFs and lognormal IMFs with a wide range of characteristic masses m_c are consistent with the data.

JWST will improve the constraints on the IMFs of nearby dwarf galaxies, and we lay out an efficient observing strategy for characterizing IMF variations in the Local Group with JWST. We summarize our results in Section 6.

2 METHODS

2.1 Lognormal IMF

We assume that the true IMF from which masses are sampled is a lognormal of the form advocated by Chabrier (2003b), which at sub-solar masses is given by

$$\xi(m) = \frac{dn}{dm} = \xi_0 \frac{1}{m} \exp \left[-\frac{(\log m - \log m_c)^2}{2\sigma^2} \right], \quad (1)$$

where the logarithm is base 10. Here $m_c = 0.22 M_\odot$ is the “characteristic mass” at which the logarithmic slope of the IMF is $d \log \xi / d \log m = -1$,¹ and $\sigma = 0.57 M_\odot$.² Hereafter, we refer to a lognormal IMF with these parameter values as a “Chabrier IMF.” The normalization factor ξ_0 is defined such that $\int_{M_{\text{min}}}^{M_{\text{max}}} \xi(m) dm = 1$, where M_{min} and M_{max} are the minimum and maximum of the mass range over which the IMF is populated. For a single-age stellar population, M_{max} is the highest initial mass which has not yet left the main sequence, and M_{min} is taken to be the hydrogen burning limit, $M_{\text{min}} \approx 0.08 M_\odot$ (e.g., Kumar 1963; Grossman 1970;

¹ Note that m_c is *not* the most probable mass or the peak of the mass probability distribution function dn/dm . The peak is at $m_c \exp\{-[\ln(10)\sigma]^2\}$, which is always less than m_c . m_c is the peak of $dn/d \log m$, the probability density function in log space.

² These parameters are for the “system” IMF, which represents the observable mass function after the single-star IMF is convolved with the distribution of unresolved multiple-star systems.

Liebert & Probst 1987). Throughout this work, we take $M_{\max} = 0.77M_{\odot}$, which is approximately the highest mass in a stellar population with $\text{Fe}/\text{H} = -2$ that remains on the main sequence after 12.5 Gyr (Paxton et al. 2011; Choi et al. 2016).

Figure 1 illustrates the typical situation for observational IMF studies of Local Group dwarf spheroidal and ultra-faint galaxies. Since these galaxies are dominated by old stellar populations, they have no surviving stars above $M_{\max} \approx 0.77M_{\odot}$, and because low-mass stars are faint, observations typically do not reach the low-mass turnover. Published observational studies of Local Group dwarf galaxies thus far have, at best, reliably sampled the IMF down to stellar masses of $M_{\text{obs}} \sim 0.4M_{\odot}$, for observations of very near satellites ($D = 60 - 70$ kpc; Wyse et al. 2002; Kalirai et al. 2013). Studies of more distant satellites (especially older studies using WFPC2) have only reached down to $M_{\text{obs}} = 0.5 - 0.6M_{\odot}$ (Grillmair et al. 1998; Geha et al. 2013).

If observations do not reach down to the low-mass turnover, one is faced with the underconstrained problem of inferring the form of the IMF from observations sampling a narrow mass range in which the IMF has a near constant slope. As Figure 1 illustrates, the Chabrier IMF is effectively indistinguishable from a power law ($\xi(m) \propto m^{-\alpha}$, where $\alpha = 2.35$ is the Salpeter value) with $\alpha \approx 1.55$ in the mass range $(0.4 - 0.77)M_{\odot}$. Data drawn from a Chabrier IMF in this mass range are also consistent with having been drawn from a wide range of lognormal IMFs with different characteristic masses m_c (and appropriately chosen σ values; e.g., $\sigma = 0.57M_{\odot}$ for $m_c = 0.22M_{\odot}$, or $\sigma = 0.36M_{\odot}$ for $m_c = 0.4M_{\odot}$). In general, if both m_c and σ are allowed to vary, observations which reach a lower mass limit M_{obs} and sample a lognormal IMF will be consistent with any lognormal IMF with $m_c \in [0, M_{\text{obs}}]$.

2.2 Sampling from the IMF

We draw masses from a lognormal IMF using inverse transform sampling, as described in Appendix A. In practice, observations cannot directly recover the true single-star IMF, even if observations are arbitrarily deep and can measure stellar masses with zero uncertainty. This is because a fraction of stars are in unresolved multiple-star systems (e.g., binaries) and thus appear in observations as a single star with the combined magnitude of all components, corresponding to a higher inferred mass. This effect will make the observationally inferred IMF slightly more top-heavy (i.e., steeper) than the underlying single-star IMF, though this effect only becomes significant at $m \lesssim 0.3M_{\odot}$ (Kroupa et al. 1991).

Rather than explicitly modeling unresolved multiple-star systems, which have poorly constrained binary fractions and mass ratio distributions, we draw masses from the “system” IMF reported in Chabrier (2003b). The system IMF is observationally motivated: it represents the IMF derived from the system luminosity function, which is produced by “merging” the components of each multiple-star system into a single unresolved source.³ The tests we carry out here thus

quantify the accuracy with which the observable system IMF can be recovered; additional modeling is required to translate the system IMF into a single-star IMF.

2.3 Fitting the IMF

Recovering the low-mass IMF from an observed sample of stars is straightforward compared to the high-mass IMF, because to first order, the galaxy’s star formation history (SFH) is irrelevant (Miller & Scalo 1979; Scalo 1986). Since the main-sequence lifetimes of low-mass ($\lesssim 0.75M_{\odot}$) stars exceed the age of the universe, the present-day mass function is identical to the IMF at low masses, irrespective of the galaxy’s star formation history.⁴

We experimented with two different methods for fitting the IMF to observations. The first method is idealized, with the assumption that the masses of individual stars can be measured exactly. In this case, the probability of measuring a particular mass is simply the value of the IMF at that mass. Assuming masses are independent and identically distributed, the likelihood function for a set of stellar masses is simply the product of the probability of each mass. Explicitly, given a sample of masses m_i and a functional form of the IMF $\xi(m, \theta_j)$, where θ_j are free parameters, the likelihood function is

$$\mathcal{L} = p(m_i | \theta_j) = \prod_i \xi(m_i, \theta_j). \quad (2)$$

An advantage of this approach compared to the traditional method of essentially fitting a line to a histogram is that it does not require any binning. Binning masses when fitting a distribution introduces unnecessary ambiguity in the choice of bins and can lead to biases in the inferred distribution, particularly when there are unequal numbers of samples in each bin (e.g. Maíz Apellániz & Úbeda 2005; Maschberger & Kroupa 2009).

We also experimented with a more realistic (though still idealized), observationally motivated fitting approach using synthetic CMDs generated for a stellar population with a realistic completeness and metallicity distribution function. In this case, it is not possible to determine the mass of any individual star exactly, since stars of different metallicities have different magnitudes at fixed initial mass.

We describe this approach in detail and compare the results of CMD fitting to the idealized approach of fitting masses directly in Appendix C. We find that, if the metallicity distribution function and observational completeness are known, the two fitting approaches yield effectively identical best-fit values, covariances, and marginalized uncertainties in IMF parameters. That is, in the limit of zero photometric uncertainty and perfect stellar models, no information is lost by fitting observables as opposed to fitting masses directly

is derived by applying a single-star mass-magnitude relationship to an unresolved multiple-star system.

⁴ Strictly speaking, SFH can have a slight effect on the recovered IMF, because there is weak evolution in the magnitudes of low-mass stars even while they remain on the main sequence. However, this effect introduces negligible uncertainty in the recovered low-mass IMF (Geha et al. 2013), because the main-sequence lifetimes of low-mass stars are many times the age of the Universe.

³ Note that this is defined differently from e.g. the system IMF of Kroupa et al. (2013), which describes the distribution of the total masses of multiple-star systems. The Chabrier system IMF

as in Equation 2. Having demonstrated as a proof of concept that both fitting approaches produce the same results in Appendix C, we use the first approach (directly fitting masses) exclusively for our primary analysis in the interest of computational cost.

For both fitting approaches, we assume flat priors for the free parameters of the IMF. When fitting a lognormal, we use the priors $m_c \in [0.01, 1.5] \times M_\odot$ and $\sigma \in [0.1, 2] \times M_\odot$; when fitting a power law, we take $\alpha \in [0, 5]$. We have verified that our priors are wide enough to be noninformative; that is, increasing the range of the priors does not change the shape of the recovered posterior.

We use the affine-invariant ensemble sampler **emcee** (Foreman-Mackey et al. 2013) to sample from the posterior. To verify that our chains have converged, we use the Gelman-Rubin convergence diagnostic (Gelman & Rubin 1991; Gelman et al. 2013). Formally, we require $\hat{R} < 1.05$ for all chains, where \hat{R} is the potential scale reduction factor (PSRF).⁵ We find that in practice, drawing 20,000 samples from the posterior is always sufficient to satisfy this diagnostic threshold.

3 RECOVERING THE PARAMETERS OF A LOGNORMAL IMF

In this section, we investigate how accurately the parameters of a lognormal IMF can be recovered from a sample of masses. We begin by exploring how the observational mass lower-limit M_{obs} and total stellar mass M_{star} of a sample affect the strength of constraints on m_c and σ , as well as the degeneracy between the two parameters. We assume that the true IMF is a Chabrier lognormal IMF with $m_c = 0.22M_\odot$ and $\sigma = 0.57M_\odot$. We consider the case in which the true IMF *does* vary in Section 4.

3.1 Overview

We simulate observations of a galaxy with total stellar mass M_{star} and a Chabrier IMF between $M_{\text{min}} = 0.08M_\odot$ and $M_{\text{max}} = 0.77M_\odot$ by drawing N_{obs} masses from the IMF between M_{obs} and M_{max} .⁶ Due to incompleteness at lower masses, only the fraction of stars c_{obs} with masses greater than M_{obs} can be observed, so

$$N_{\text{obs}} = N_{\text{tot}} c_{\text{obs}} = \frac{M_{\text{star}}}{\bar{m}} c_{\text{obs}} = M_{\text{star}} \frac{\int_{M_{\text{obs}}}^{M_{\text{max}}} \xi(m) dm}{\int_{M_{\text{min}}}^{M_{\text{max}}} \xi(m) m dm}. \quad (3)$$

In practice, the lower-limit for observations is not a threshold mass, but rather a magnitude threshold below

⁵ The PSRF quantifies the ratio of the mean variance within individual chains to the variance of the mean across all chains. In the limit of infinite samples, $\hat{R} \rightarrow 1$; values of \hat{R} greater than ~ 1.1 indicate poor convergence (Gelman et al. 2013).

⁶ In this formulation, M_{star} represents the total stellar mass in the region of the galaxy covered by observations. If only a fraction of the galaxy is observed, the total stellar mass will be greater than M_{star} . Note also that M_{star} represents only the initial mass contribution of main-sequence stars. We do not account for mass loss, evolved stars, or remnants.

which completeness falls off steeply because stars are too faint to observe or observations become crowding limited. We translate magnitude thresholds in a given filter to M_{obs} by interpolating on a grid of isochrones from the MESA Isochrones and Stellar Tracks project (MIST; Choi et al. 2016), as explained in Appendix B. For this conversion, we assume a stellar population of age 12.5 Gyr with metallicity $[\text{Fe}/\text{H}] = -2$.

3.2 Fitting results

Figure 2 shows the constraints on m_c and σ obtained by fitting a lognormal IMF to a range of mass samples drawn from a Chabrier IMF. The white lines and blue shaded regions respectively show the median and middle 68 percent of the marginalized posterior distributions for m_c (top) and σ (bottom). Large shaded regions indicate significant uncertainty in the IMF parameters, due both to the degeneracy between m_c and σ and to stochastic sampling effects arising from the finite number of stars being sampled (Elmegreen 1999; Hernandez 2012). Smaller regions indicate that m_c and σ are individually well constrained.

In the left two panels, we generate a stellar population with total mass $M_{\text{star}} = 10^4 M_\odot$ (corresponding to $N_{\text{tot}} = 34,287$ between $0.08M_\odot$ and $0.77M_\odot$) and vary M_{obs} , the mass down to which stars can be observed. At $M_{\text{obs}} = 0.4M_\odot$, this M_{star} corresponds to $N_{\text{obs}} = 8,686$ stars included in the fit, which is a factor of a few more than the number used in most previous observational studies of Local Group dwarf galaxies.

For $M_{\text{obs}} \gtrsim 0.2M_\odot$, the middle 68% ranges for both m_c and σ are large, indicating weak constraints on the individual IMF parameters. As we will show, this is largely the result of the degeneracy between m_c and σ . Strong constraints on m_c and σ can be obtained only when $M_{\text{obs}} \lesssim m_c$. This makes intuitive sense, since the lognormal IMF has a nearly constant slope at $m \gg m_c$ and only begins to turn over at $m \sim m_c$. Above the turnover, m_c and σ are strongly degenerate, because a wide range of lognormal IMFs with $0 \lesssim m_c \lesssim M_{\text{obs}}$ can all fit the data equally well (see Figure 1).

In the right panels of Figure 2, we again show constraints on m_c and σ , but now vary M_{star} while $M_{\text{obs}} = 0.4M_\odot$ is held fixed. As expected, stronger constraints can also be obtained by increasing M_{star} , but very large numbers of stars are required to obtain meaningful constraints. For example, M_{star} must be increased by a factor of ~ 10 to obtain the same improvement on the fiducial constraints that can be obtained by decreasing M_{obs} by $0.2M_\odot$.

For reference, Figure 2 also shows the absolute magnitudes of individual stars (left) and the total galaxy (right) that correspond to the plotted values of M_{obs} and M_{star} . We calculate M_V for individual stars from the MIST isochrones as explained in Appendix B and M_V for a whole galaxy using Equation 4.

3.3 Fitting a power law IMF to data drawn from a lognormal

In the absence of sufficiently deep observations with the statistical power to jointly constrain m_c and σ of a lognormal

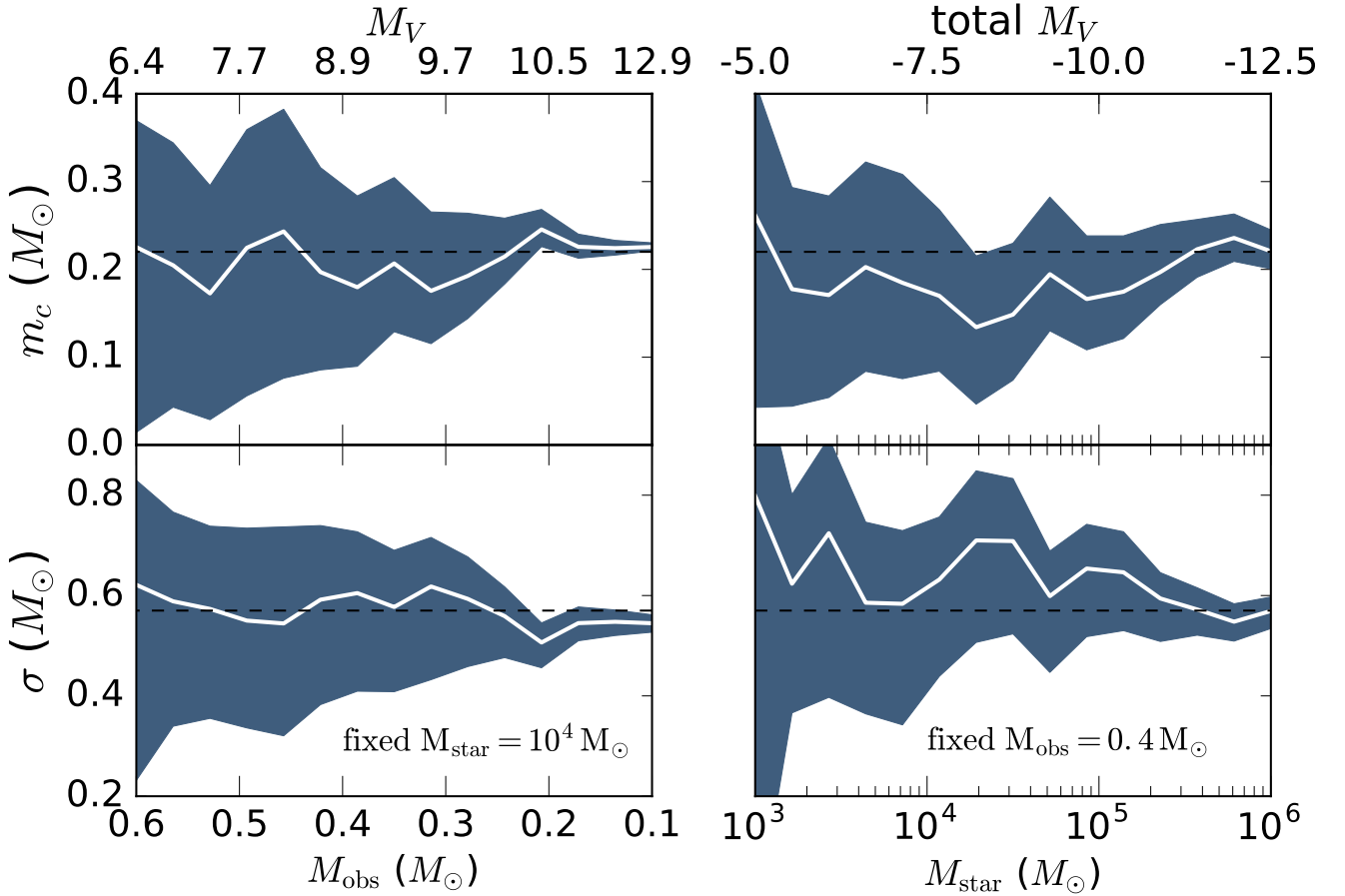


Figure 2. Best-fit values of m_c (top) and σ (bottom) obtained from fitting a lognormal IMF to samples of stars drawn from a Chabrier IMF. White line and shaded regions show the median and middle 68% of the marginalized posterior for each IMF parameter. **Left:** we fix the total mass of all stars in the range $0.08 - 0.77 M_\odot$ at $10^4 M_\odot$ and vary M_{obs} , the mass threshold down to which stars can be observed. **Right:** we fix M_{obs} at $0.4 M_\odot$ (corresponding to a magnitude threshold of $M_V \approx 8.9$) and vary the total stellar mass. Tight constraints on m_c and σ can be obtained either by observing systems with more stars or by pushing observations to fainter magnitudes, but unrealistically large stellar samples are needed to meaningfully constrain the low-mass IMF at the magnitude limits of current studies with HST.

IMF, it is common to make simplifying assumptions, such as fitting a single power law in place of a lognormal or holding σ fixed when fitting a lognormal (e.g. Grillmair et al. 1998; Wyse et al. 2002; Geha et al. 2013; Kalirai et al. 2013). This avoids the degeneracy between m_c and σ , since either approach essentially reduces the fit to a one dimensional optimization problem.

Figure 3 shows the results of fitting a power law ($\xi(m) \propto m^{-\alpha}$) to masses drawn from a Chabrier IMF. As in the left panel of Figure 2, we hold M_{star} fixed at $10^4 M_\odot$ while varying M_{obs} . Here, the 68 percent probability range in α is narrow: for $M_{\text{obs}} \lesssim 0.5 M_\odot$, α is constrained within $\lesssim 10\%$. The improved constraints on α at lower M_{obs} primarily reflect the fact that deeper observations increase the total number of stars included in the fit.

The best-fit logarithmic slope α becomes shallower for decreasing M_{obs} , since deeper observations begin to sample the mass regime near the turnover in the IMF. However, for values of M_{obs} in the range $(0.3 - 0.6) M_\odot$ (with $M_{\text{max}} = 0.77 M_\odot$), the best-fit logarithmic slope is almost constant at

$\alpha \approx 1.55$. This suggests that, for this choice of $M_{\text{star}} = 10^4 M_\odot$, observations would have to reach $M_{\text{obs}} \lesssim 0.3 M_\odot$ before they could begin to detect deviations from a power law.

Figure 3 also shows the results of previous studies which have fit power law IMFs to star count data from Local Group dwarf galaxies. Black stars show results for ultra-faint and dwarf spheroidal galaxies, which have very low metallicities, no gas, and are dominated by old stellar populations; we plot results for Draco (Grillmair et al. 1998), Ursa Minor (Wyse et al. 2002), Hercules and Leo IV (Geha et al. 2013). We also show the result of Kalirai et al. (2013) for the IMF of the SMC; we plot this point in blue to emphasize the SMC’s more complicated star formation history. These previous studies have reached down to M_{obs} values between $0.37 M_\odot$ and $0.6 M_\odot$. None of them detected a turnover at lower masses, and they thus all opted to fit a single power law.

We note that the best-fit logarithmic slope of $\alpha \approx 1.55$ for masses in the range typically probed by observations ($0.4 M_\odot \lesssim m \lesssim 0.77 M_\odot$) is significantly shallower than the

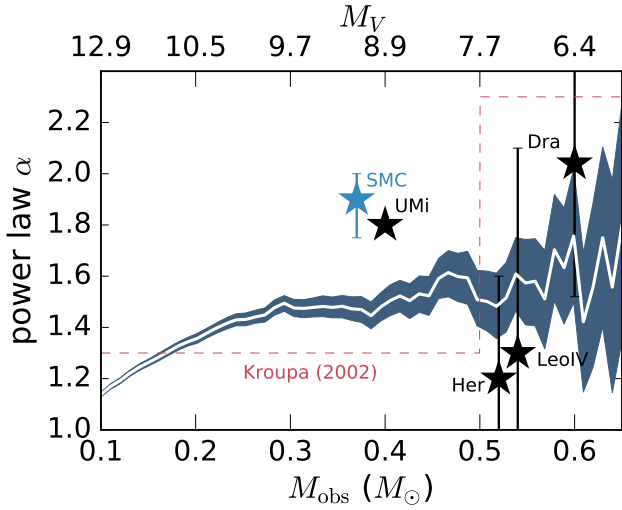


Figure 3. Best fit logarithmic IMF slope ($\xi(m) \propto m^{-\alpha}$) obtained by fitting a power law to masses drawn from a Chabrier lognormal IMF (with total $M_{\text{star}} = 10^4 M_{\odot}$) between M_{obs} and $M_{\text{max}} = 0.77 M_{\odot}$. Fitting a power law to observations which reach $M_{\text{obs}} \sim 0.5 M_{\odot}$ yields a logarithmic slope $\alpha \sim 1.55$, which is significantly shallower than the canonical value of $\alpha = 2.3$. Observationally-inferred IMF constraints are overplotted for five Local Group dwarf galaxies: Ursa Minor (Wyse et al. 2002, no error bars included), Hercules and Leo IV (Geha et al. 2013), Draco (Grillmair et al. 1998), and the Small Magellanic Cloud (SMC; Kalirai et al. 2013). Note that the first four galaxies are dominated by old stellar populations, while the SMC sample contained some younger stars.

Kroupa (2002) value of $\alpha = 2.3$ for $m > 0.5 M_{\odot}$ and the Salpeter (1955) value of $\alpha = 2.35$ at all masses. Previous studies have suggested that the shallower α values measured in Local Group dwarf galaxies are evidence of systemic variation in the IMF with environment. For example, Geha et al. (2013) pointed out that the value $\alpha = 1.2^{+0.4}_{-0.5}$ found in Hercules is shallower than the Kroupa value at the 5.4σ level. However, Figure 3 shows that the shallowest observationally-inferred IMFs are all approximately consistent with having been drawn from a Chabrier lognormal IMF.

There are two reasons for this somewhat counterintuitive result. First, the Chabrier and Kroupa IMFs, though similar in overall shape, have different instantaneous slopes: at $m = 0.5 M_{\odot}$, the Chabrier IMF has an instantaneous logarithmic slope $d \log \xi / d \log m \approx -1.48$; at the same mass, the logarithmic slope of the Kroupa IMF steepens to -2.3 .⁷ Thus, slopes of $\alpha \sim 1.55$ at $m \geq 0.5 M_{\odot}$ appear significantly shallower than the Kroupa value (the 99% (3σ) confidence interval on the Kroupa slope is $\alpha = 2.3 \pm 0.3$ in this mass range (Kroupa 2001)), but are entirely consistent with being drawn from a Chabrier IMF. We therefore stress that, although many works treat both the Kroupa and Chabrier IMFs as identical “canonical” IMFs, the two IMFs make sig-

nificantly different predictions for observations which sample only a narrow range of masses near $0.5 M_{\odot}$.

Second, the logarithmic slope of the Kroupa IMF steepens discontinuously from $\alpha = 1.3$ to $\alpha = 2.3$ at $m = 0.5 M_{\odot}$. Thus, although the shallower IMF slopes found in Local Group dwarf galaxies thus far *are* significantly different from the Kroupa value at $m \geq 0.5 M_{\odot}$, they are consistent with the Kroupa value at slightly lower masses. As there is little physical reason to expect a discontinuity in the slope of the IMF (Miller & Scalo 1979), one might reasonably expect intermediate α values between 1.3 and 2.3 at masses near $0.5 M_{\odot}$.

Kalirai et al. (2013) interpreted their best-fit slope of $\alpha = 1.90^{+0.15}_{-0.10}$ as an indication that the SMC’s IMF is somewhat shallower at lower masses than the Kroupa value, and significantly shallower than the $\alpha = 2.35$ prediction from Salpeter (1955). Figure 3 shows that the measured slope is actually somewhat *steeper* than what would be measured if masses were drawn from a Chabrier IMF in the range $0.37 M_{\odot} \leq m \leq 0.77 M_{\odot}$. However, Kalirai et al. (2013) fit masses up to $M_{\text{max}} = 0.93 M_{\odot}$ rather than $0.77 M_{\odot}$. Adopting their M_{max} and M_{min} , we find $\alpha = 1.58^{+0.06}_{-0.05}$ for masses drawn from a Chabrier IMF. This is slope is shallower than the SMC value, but only at the $(2-3)\sigma$ level.

Finally, we note that, with the exception of the data for Ursa Minor, the observational results plotted in Figure 3 were obtained in studies that attempted to explicitly model the effects of binaries and recover the single-star IMF, not the system IMF. The system IMFs corresponding to these results are expected to be slightly shallower (lower α). However, as demonstrated by Kroupa et al. (1991, their Figure 3), accounting for binarity has negligible effects on the inferred IMF except at very low masses ($m \lesssim 0.3$). We find that at $m \geq 0.4 M_{\odot}$, the single-star and system IMFs reported in Chabrier (2003a) differ by less than 10%, so we do not expect binarity to significantly change our results.

However, future observations with JWST will probe the IMFs of nearby dwarf galaxies to significantly lower masses (see Section 5), and we expect more significant differences between the system and single-star IMFs in this regime. In particular, because the conversion between the system and single-star IMFs depends on the assumed binary fraction, environmental variations in the binary fraction or mass ratios in binary systems could lead to variation in the system IMF, even in the absence of variation in the single-star IMF.

3.4 Distinguishing a lognormal IMF from a power law

We now consider how deep observations must reach in order to robustly detect deviations from a power law if the true IMF is lognormal. Existing observational studies of Local Group dwarf galaxies have shown that their IMFs do not show any deviation from a power law within the mass range of observations, so we wish to determine whether this constitutes evidence that these galaxies have IMFs different from the MW.

To this end, we draw masses from a Chabrier lognormal IMF and fit them with both a power law and a lognormal IMF, for observations reaching a range of values of M_{obs} and M_{star} . Then, for each set of model fits, we use the Bayesian Information Criterion (BIC; Schwarz 1978) to evaluate whether the sample of masses is fit significantly bet-

⁷ Note, however, that it is not the instantaneous slope at M_{obs} , but rather the average slope between M_{obs} and M_{max} , that drives the fit.

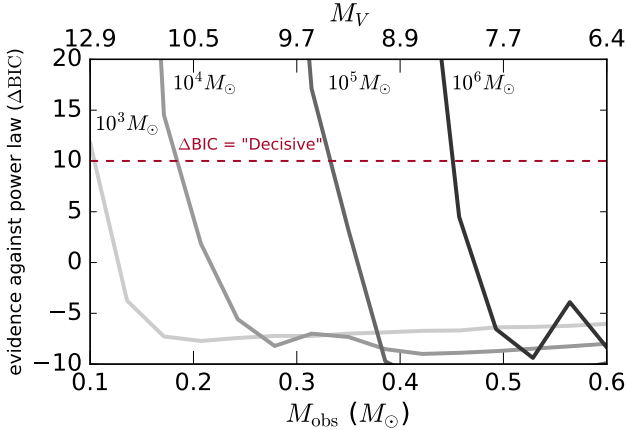


Figure 4. Difference in Bayesian information criterion (BIC) between a power law IMF and a lognormal. Higher ΔBIC indicates stronger deviations from a power law and evidence for a lognormal. In all cases, we draw masses from a Chabrier lognormal IMF. Then, using only masses above a threshold M_{obs} , we fit both a power law and a lognormal to the sampled masses and calculate the BIC. Correctly inferring that the IMF is lognormal, not a power law, requires either deep observations (low M_{obs}) or a very large sample (high M_{star}).

ter by a lognormal IMF than by a power law. The BIC is a rough diagnostic for comparing how well different models fit a particular dataset. It is designed to balance higher likelihood against increasing model complexity, since more complex models with larger numbers of free parameters can often produce higher likelihood through overfitting.

In general, models with lower BIC are preferred. In Figure 4, we plot the difference in BIC between a power law and lognormal IMF as a function of M_{obs} for different values of M_{star} . A common rule of thumb is that a $\Delta\text{BIC} > 10$ is considered “decisive” evidence against the model with higher BIC (Liddle 2007); thus, IMF realizations which lie above the dashed line in Figure 4 strongly favor a lognormal IMF over a power law. Points below the line indicate that there is not a strong preference for one model or the other, or (for $\Delta\text{BIC} < -10$), that a power law model is strongly preferred over a lognormal.

Figure 4 shows that for $M_{\text{star}} = 10^4 M_{\odot}$, the lognormal fit is not decisively preferred until $M_{\text{obs}} \sim 0.2$, comparable to m_c . This can be understood intuitively: the lognormal IMF has a nearly constant slope at $m \gg m_c$, and it is thus not possible to distinguish between a simple power law and an IMF which turns over at lower masses with insufficiently deep data. However, statistical power to constrain the IMF increases with increasing M_{star} , so that at fixed M_{obs} , deviations from a power law IMF can be detected more robustly for larger M_{star} .

We note that none of the published IMF constraints shown in Figure 3 reach sufficiently low masses to robustly rule out a turnover at the Chabrier characteristic mass. The constraints from Kalirai et al. (2013) come closest, but we find that for their sample size (~ 5000 stars with masses below $0.93 M_{\odot}$), observations would have to reach down to $\sim 0.25 M_{\odot}$ before the BIC would decisively favor a lognormal to a power law IMF.

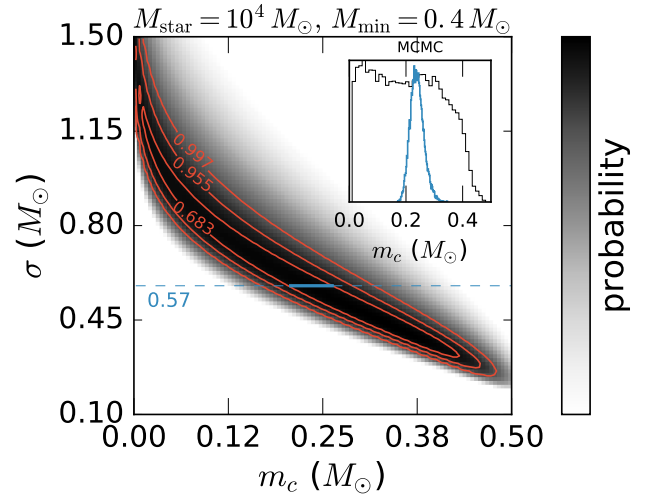


Figure 5. Likelihood for fitting a lognormal IMF. Orange contours enclose regions corresponding to 68.3%, 95.4%, and 99.7% probability. Inset shows samples from the posterior distribution for m_c obtained from MCMC fitting. Blue histogram corresponds to holding σ fixed; black histogram shows results of allowing σ to vary and marginalizing. Because m_c and σ are highly degenerate, holding σ fixed during fitting significantly underestimates the uncertainty in m_c .

In conclusion, we have shown that published star count data in the Local Group do not reach sufficiently low stellar masses to robustly measure the low-mass turnover in the IMF: existing studies have only measured the slope of the IMF over a narrow mass range ($0.4 M_{\odot} \lesssim m \lesssim 0.8 M_{\odot}$) in which a lognormal IMF is effectively indistinguishable from a power law. If the IMF is lognormal, observations must reach approximately down to m_c , which for a Chabrier IMF is $m \approx 0.22$ ($M_V \approx 10.2$ for an old, metal-poor stellar population).

In addition, we find that the shallower IMF slopes reported in Local Group dwarf galaxies thus far between masses of $\sim 0.4 M_{\odot}$ and $\sim 0.8 M_{\odot}$ are approximately consistent with a Chabrier IMF, despite apparent disagreement with the Kroupa slope in this mass regime. We emphasize that deeper observations are required for robust detection of IMF variations.

3.5 Fixing model parameters during fitting

Instead of fitting a power law IMF, another possible approach for constraining the IMF when data are only available over a narrow range of masses is to hold one of the IMF parameters fixed during fitting. If either m_c or σ is held fixed (and $m_c \ll M_{\text{obs}}$, so the IMF resembles a power law), then varying the parameter which is not held fixed effectively varies the power law slope at $m \gg m_c$.

Figure 5 shows isoproability contours for fitting a lognormal IMF to a sample of masses and compares the constraints on m_c obtained when σ is held fixed at $\sigma = 0.57 M_{\odot}$ (the Chabrier value) to those obtained when both σ and m_c are allowed to vary. In the inset, we show draws from the marginalized posterior for m_c when both parameters are al-

lowed to vary (black) and when σ is held fixed at $0.57M_{\odot}$ (blue).

When both parameters are left free, m_c and σ are highly covariant. Although essentially any $m_c \lesssim 0.4M_{\odot}$ provides a good fit when paired with an appropriate choice of σ , all acceptable fits reside in a nearly one-dimensional region of parameter space. Thus, when σ is held fixed, fitting a lognormal yields much stronger constraints on m_c .

It is important to note, however, that holding σ fixed essentially amounts to putting a delta function prior on σ , which is generally not justified. Because of the strong degeneracy between m_c and σ , holding σ fixed will dramatically underestimate the true uncertainty in m_c : naively interpreting the blue posterior for m_c in Figure 5, one might conclude that the characteristic mass m_c is well-constrained to $0.2M_{\odot} < m_c < 0.3M_{\odot}$ when in fact, a much larger range of values can provide equally good fits. This is not necessarily a problem, *if* the constraints from fitting are compared to other studies in the same mass range, since all acceptable fits have basically the same shape across the mass range where data are available (see Figure 1).

On the other hand, attempts to extrapolate the fitted IMF to lower masses in order to constrain the turnover mass will yield erroneously strong constraints: clearly, if observations only sample a small range of masses at $m \gg m_c$, where the IMF has not yet begun to turn over, they cannot reliably constrain m_c or the turnover mass. We therefore caution against fixing model parameters during fitting. If observations do not reach sufficiently low-mass stars to measure the turnover in the IMF, it is better to fit a simple power law than to fit a more complicated IMF while fixing one or more of the parameters.

4 WHAT IS NEEDED TO DETECT VARIATION IN THE IMF?

One of the goals of studying the IMFs of dwarf spheroidal and ultra-faint galaxies is to detect or rule out systematic IMF variations with environment. In this section, we investigate how deep observations must reach in order to robustly detect variation in the IMF (if it exists). We assume that the IMF is lognormal, but that the characteristic mass m_c varies across galaxies. Since observations of nearby dwarfs have found only marginal evidence of IMF variations at $m \gtrsim 0.5M_{\odot}$, we intentionally choose m_c and σ so that all IMFs have approximately the same slope at $m \sim 0.5M_{\odot}$.

Figure 6 shows how accurately the form of the IMF can be recovered for different M_{obs} , the minimum mass down to which observations are complete. For each IMF sample realization, we plot the median and middle 68% percent range of the marginalized posterior for m_c against the true m_c of the IMF from which the sample was drawn. If the middle 68% region contains $m_c = 0.22M_{\odot}$, we label the IMF realization “consistent with Chabrier”; i.e., observing such a sample of masses would not raise any suspicion of variation in the IMF. If the Chabrier value of $m_c = 0.22M_{\odot}$ falls outside the middle 68% credibility region, we label the realization inconsistent with the Chabrier value. Of course, if the middle 68% credi-

bility region extends almost to $m_c = 0.22M_{\odot}$, this would not represent strong evidence of variation in the IMF.⁸

Figure 6 shows that variation in the IMF can only be robustly detected if $M_{\text{obs}} \sim m_c$; that is, if observations are deep enough to detect the beginning of the turnover. In the right panel, where $M_{\text{obs}} = 0.5M_{\odot}$, a wide range of IMFs with $0.1M_{\odot} \lesssim m_c \lesssim 0.5M_{\odot}$ are all consistent with the Chabrier value. In other words, if observations only reach down to $0.5M_{\odot}$, one could observe galaxies with lognormal IMFs and characteristic masses anywhere in the range $0.1M_{\odot} \lesssim m_c \lesssim 0.5M_{\odot}$, and the data would not provide sufficiently strong constraints to rule out a Chabrier IMF with $m_c = 0.22M_{\odot}$.

However, it becomes much easier to accurately constrain IMF parameters, and thus, to detect deviations from a Chabrier IMF, for lower M_{obs} . For $M_{\text{obs}} = 0.3M_{\odot}$ (middle), only the IMFs with $0.2M_{\odot} \lesssim m_c \lesssim 0.4M_{\odot}$ are approximately consistent with the Chabrier value, and typical 68% error bars are smaller than for $M_{\text{obs}} = 0.5M_{\odot}$. Finally, for $M_{\text{obs}} = 0.1M_{\odot}$, m_c can be constrained with percent-level accuracy in all cases, and all the IMFs with m_c different from the Chabrier value can be ruled out. The dramatically improved constraints on m_c for lower M_{obs} in this (admittedly idealized) experiment are the result of two factors: first, observations reach low enough that the turnover can be detected directly for lower M_{obs} , and second, the total number of stars contributing to the fit increases for deeper observations.

We note that, even for the cases with higher M_{obs} , it is always possible to obtain reasonably tight constraints on m_c (and to detect deviations compared to the Chabrier value) when $M_{\text{obs}} \lesssim m_c$. This is because when $M_{\text{obs}} \lesssim m_c$, the low-mass turnover is directly probed by observations, and the IMF is no longer well-fit by a power law over the mass range where data are available. Thus, there is no longer a significant degeneracy between m_c and σ , and both parameters can be constrained strongly. Existing observations can thus already rule out lognormal IMFs with $m_c \gtrsim 0.4M_{\odot}$.

5 LOOKING FORWARD TO JWST

Having quantified the sample size and limiting magnitude required to constrain generic lognormal IMFs using star counts, we now turn to the prospect of measuring the IMFs of Local Group dwarf galaxies with JWST. How accurately the IMF of an individual galaxy can be constrained depends on M_{obs} (which is a function of the magnitude down to which observations are complete) and on the total number of stars sampled (which is a function of the angular size of the galaxy and the number of pointings).

We begin by calculating a realistic limiting magnitude and corresponding M_{obs} for observations with JWST. Because cool, low-mass stars are most luminous in the NIR (for

⁸ We also experimented with using a KS test as a non-parametric measure of whether a given sample of masses was consistent with being drawn from a Chabrier IMF, finding results that were similar with the metric described above. We opted to use the parametric version for our primary analysis because we found significant variations in the KS-test p values across different realizations of draws from identical IMFs. This is a generic shortcoming of the KS test; see Ivezić et al. (2014).

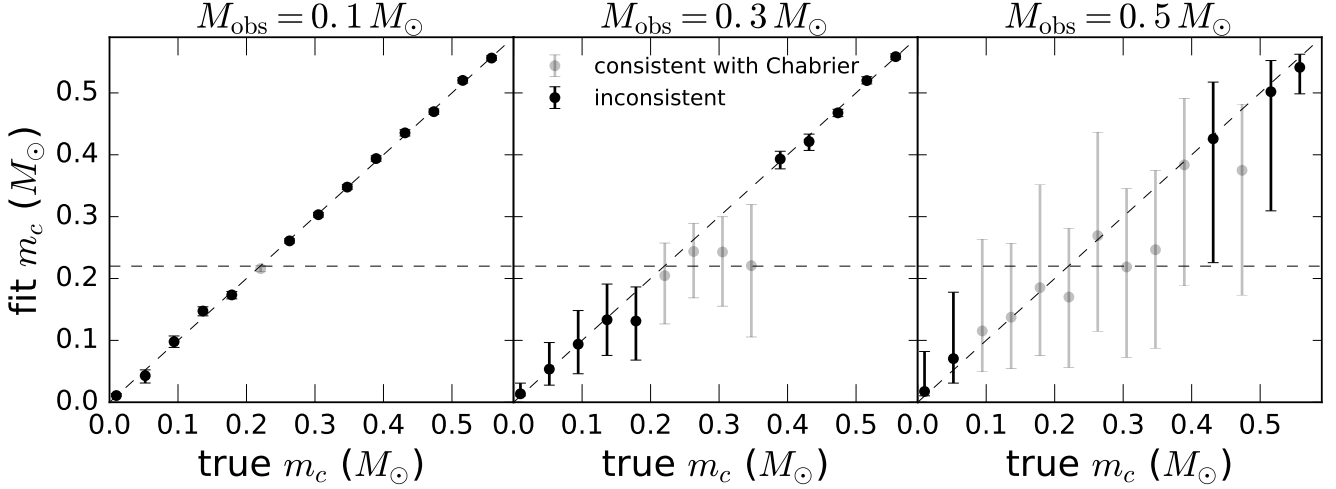


Figure 6. Accuracy to which the characteristic mass m_c of a lognormal IMF can be recovered for three different observational stellar mass lower limits, M_{obs} . We draw masses from 14 different lognormal IMFs with characteristic masses m_c in the range $0.01M_{\odot} - 0.6M_{\odot}$, always for a total $M_{\text{star}} = 10^4 M_{\odot}$. For each m_c , we choose σ such that the IMF at masses above $0.5M_{\odot}$ resembles a Chabrier IMF (see Figure 1). We then fit a lognormal IMF to the drawn masses. Points and error bars show the median and middle 68% of the marginalized posterior draws for m_c . Gray points indicate that the data cannot rule out a Chabrier IMF with $m_c = 0.22M_{\odot}$; black points indicate that $m_c = 0.22M_{\odot}$ falls outside the 68% constraints on m_c , and thus, that the data appear to be inconsistent with being drawn from a Chabrier IMF.

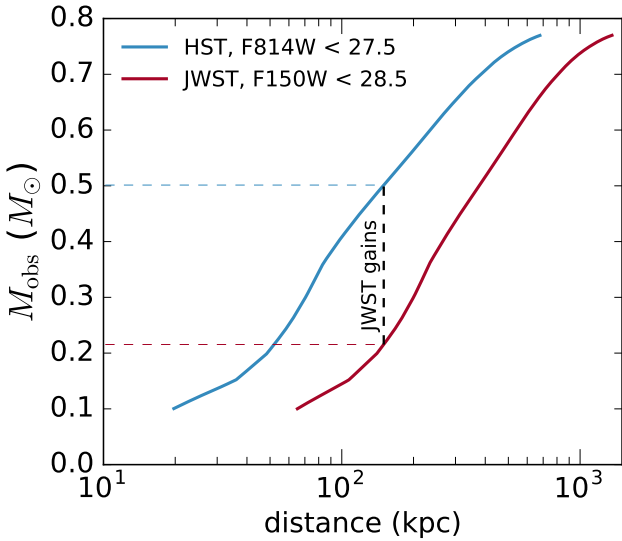


Figure 7. Lowest observable stellar mass as a function of distance for HST ACS and JWST observations, assuming 20,000 seconds of integration time. At a fixed distance, JWST can reach stars nearly $0.3M_{\odot}$ less massive than HST. JWST observations will reach the hydrogen burning limit for distances $\lesssim 60$ kpc.

a typical low-metallicity M-dwarf, f_{λ} peaks near $0.9\mu\text{m}$), it is best to use filters near this wavelength. We use the STScI JWST exposure time calculator⁹ to determine the optimal filters for observing low-mass stars and the projected limiting magnitude in these filters. We find that the *F150W* and

F200W filters are optimal for observations of low-mass stars, because although the peak in f_{λ} is somewhat bluer, these filters are significantly wider than the *F090W* and *F115W* filters and thus admit more total photons. We find that for an exposure time of approximately 20,000 seconds in the *F150W* filter, JWST can obtain a signal-to-noise ratio $\text{SNR} = 5 - 10$ or better for stars brighter than 28.5 Vega mag.

Figure 7 shows the expected lowest mass M_{obs} which JWST will reach as a function of distance. We compute M_{obs} from apparent magnitude at a given distance by interpolating on a grid of isochrones using the MIST code. For comparison, we show the approximate M_{obs} limit of current optical observations with HST.¹⁰ This HST curve is computed assuming a magnitude limit of 27.5 Vega mag in the *F814W* filter, which can be obtained for an exposure time and desired S/N ratio similar to our assumptions for JWST (Geha et al. 2013; Kalirai et al. 2013).

Promisingly, JWST is forecasted to reach significantly lower mass stars at fixed distance. For example, at $d = 150$ kpc, the approximate distance of the most-distant MW satellites for which previous studies have attempted to measure the IMF (Geha et al. 2013), HST observations can only reach down to $M_{\text{obs}} \approx 0.5M_{\odot}$ with typical integration times, which, as we showed in Section 3.2, is not deep enough to jointly constrain σ and m_c for a Chabrier IMF. At the same distance, JWST is projected to reach down to $M_{\text{obs}} \sim 0.22M_{\odot}$,

¹⁰ Previous studies attempting to constrain the IMFs of nearby dwarf galaxies have used the HST *F606W* and *F814W* filters rather than NIR filters, both because HST is significantly less sensitive (and has poorer angular resolution) in the NIR than at optical wavelengths, and because the low-mass stars reached by previous studies ($m \sim 0.5M_{\odot}$) are somewhat less red than the lowest-mass stars that can be reached with JWST.

⁹ <https://jwst.etc.stsci.edu/>

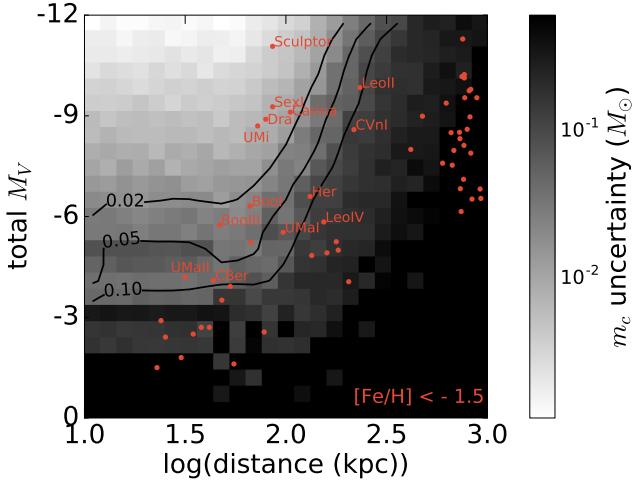


Figure 8. Projected accuracy to which the IMF can be recovered from JWST observations of Local Group dwarf galaxies. Gray pixels show the strength of constraints on m_c obtained by fitting a lognormal IMF to samples of stars drawn for populations with different M_{obs} and total M_V ; lighter pixel colors indicate stronger constraints. Local Group dwarf galaxies with $[\text{Fe}/\text{H}] < -1.5$ are overplotted, with the most promising targets labeled.

which is sufficiently deep to detect the low-mass turnover and constrain m_c and σ within $\sim 20\%$, if the IMF is similar to a Chabrier IMF. In addition, for the nearest MW satellites at $d \lesssim 60$ kpc, JWST observations will reach stars down to the hydrogen burning limit. These significant improvements over HST stem both from JWST’s larger collecting area and from its sensitivity to longer wavelengths, where low-mass stars are more luminous.

5.1 Optimal targets in the Local Group

We now combine the projected sensitivity of JWST with the dependence of IMF constraints on M_{star} and M_{obs} derived in the previous sections to devise an observational strategy for constraining the low-mass IMF in the Local Group. Nearer galaxies can be observed down to fainter M_{obs} , but there are more massive galaxies, which offer better statistical power (i.e., larger M_{star}), at larger distances. An additional consideration is that more distant objects subtend a smaller angular size on the sky, and thus require fewer pointings, but are also more likely to be affected by crowding.

We investigate the tradeoff between distance and the number of observable stars in Figure 8, which shows the constraints on m_c expected for a galaxy at a given distance (which determines M_{obs}) with a given absolute magnitude (which determines M_{star}). For a given total M_V , the total stellar mass is

$$M_{\text{star}} = f_{\text{obs}} \Upsilon_V \times 10^{-M_V/2.5}. \quad (4)$$

Here Υ_V is the V-band stellar mass-to-light ratio:¹¹

$$\Upsilon_V = \frac{\int_{M_{\text{min}}}^{M_{\text{max}}} \xi(m) m dm}{\int_{M_{\text{min}}}^{M_{\text{max}}} \xi(m) L_V(m) dm}, \quad (5)$$

where $L_V(m)$ is the V-band luminosity of a star with initial mass m , normalized at the magnitude zeropoint. In practice, we calculate $L_V(m)$ for a population of a given age and metallicity by interpolating on a grid of stellar models. We assume that observations will only observe a fraction $f_{\text{obs}} = 0.1$ of all stars above M_{obs} , due to the limited number of pointings and the difficulty of distinguishing between the satellite and foreground MW stars in the galaxy’s outskirts. That is, f_{obs} represents the fraction of a galaxy’s angular area that is imaged, and the observable $M_{\text{star}} = f_{\text{obs}} M_{\text{star,tot}}$.

In each pixel of Figure 8, we calculate the value of M_{star} corresponding to M_V using Equation 4, and the value of M_{obs} by interpolating the JWST curve in Figure 7. We then draw masses between M_{obs} and M_{max} from a Chabrier IMF of total mass M_{star} between M_{min} and M_{max} , following the procedure outlined in Section 3. Finally, we fit a lognormal IMF to the drawn masses using our standard MCMC procedure outlined in Section 2.3. We measure the 68% uncertainty in the recovered value of m_c and color each pixel in Figure 8 accordingly. Lightly shaded pixels indicate that m_c could be recovered with little uncertainty, while dark pixels correspond to large uncertainties and poor constraints on m_c .

In Figure 8, we also overplot Local Group dwarf galaxies that represent possible targets for observational studies to measure the IMF from direct star counts. Data for these galaxies are taken from the Nearby Dwarfs Database presented in McConnachie (2012). We only plot low-metallicity objects with measured $[\text{Fe}/\text{H}] < -1.5$. These represent the most “extreme” environments, in which one might most expect systematic IMF variations compared to the MW. In addition, these galaxies typically have primarily old stellar populations with simple SFHs, which makes it easier to determine their IMFs from a CMD. “Promising” objects (i.e., those in regions of parameter space where our simulations suggest that the IMF could be constrained with little uncertainty) are labeled. Points in the upper right corner are primarily the satellites of M31, which are too faint and crowded to be promising targets.

So far, we have only considered the M_{obs} and M_{star} that can be achieved for targets of a given M_V at a given distance, without taking into account possible complications due to crowding or the number of pointings required for a given source. We now consider the effects of these issues, in order to refine the list of promising targets.

In Figure 9, we plot the half-light radius, r_h , and maximum central surface number density of stars, Σ_0 , of the same galaxies shown in Figure 8. We roughly estimate Σ_0

¹¹ Note that unlike in the previous sections, the integral is over all stars, including those that have left the main sequence. The resulting Υ_V represents a time-averaged quantity, which can fluctuate significantly due to stochastic sampling in the faintest galaxies, whose light is dominated by a small number of evolved stars (Hernandez 2012).

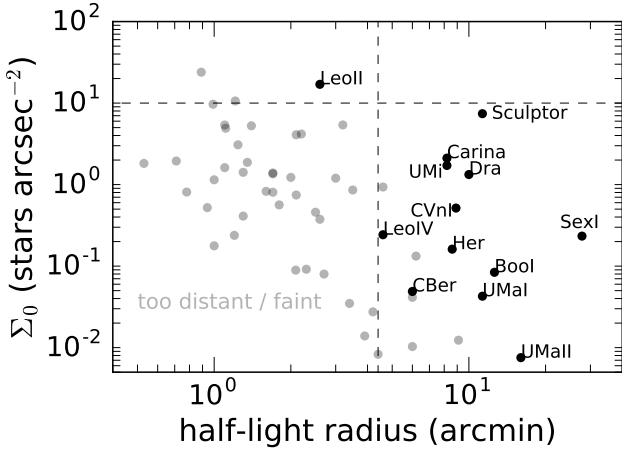


Figure 9. Maximum central stellar surface density vs. half-light radius of the same Local Group dwarf galaxies plotted in Figure 8. Gray dots represent targets found to be unfavorable in Figure 8, which are also not labeled in Figure 8. For comparison, we plot as dashed lines the angular size of JWST’s NIRCам field of view (4.4 arcmin) and the surface density above which observations will become limited by crowding ($\approx 10 \text{ stars arcsec}^{-1}$).

from M_V , r_h , and distance under the assumption that stars follow an exponential surface density profile:

$$\Sigma_0 = \frac{N_{\text{obs}}}{2\pi a^2}, \quad (6)$$

where N_{obs} is the number of stars above M_{obs} (see Equation 3), and $a = r_h/1.678$ is the exponential scale length (conversion from Wolf et al. (2010)). Note that Σ_0 represents the maximum surface density of stars above the M_{obs} reachable with a 20,000 JWST second exposure, not the total surface density of stars down to the hydrogen burning limit.

We also show the angular size of JWST’s NIRCам (4.4 arcmin; 2.2 in the other dimension), and the surface density threshold of $10 \text{ stars arcsec}^{-2}$, which is approximately the surface density above which observations become significantly crowding limited. This value is determined under the assumption that JWST’s crowding limits will be comparable to those of HST in optical filters, using the crowding data for HST from Dalcanton et al. (2012).

Ideal targets would have both low surface densities (so that crowding is not an issue) and small half-light radii (so that a small number of pointings can image the whole galaxy). Figure 9 shows that crowding should not present a significant problem for most of the promising targets, most of which have maximum stellar surface densities more than a factor of 10 below the assumed crowding threshold. On the other hand, nearly all of the promising targets are significantly larger than the JWST field of view.

We note that, since we assumed that only a fraction $f_{\text{obs}} = 0.1$ of stars would be imaged in creating Figure 8, it is not necessary to image the entire galaxy to obtain the projected constraints on m_c . Nevertheless, the majority of the targets labeled in Figure 9 would require more than one pointing. This is not altogether surprising, since constraining the IMF at low masses requires low M_{obs} , which can only be reached for nearby satellites, and nearby satellites subtend the largest angular size at fixed M_V .

One possibility for observations in the post-JWST era is to use WFIRST, which will have a field of view nearly 100 times larger than JWST, and thus could image any of our promising targets in a single pointing (Spergel et al. 2013). Unfortunately, WFIRST will have significantly poorer angular resolution than JWST, and will thus be more crowding limited. In addition, WFIRST will not reach as low an M_{obs} as JWST at fixed distance and exposure time due to its smaller collecting area. Thus, WFIRST will be ideally suited for measuring the IMFs of the very nearest MW satellites (e.g. Coma Berenices, Ursa Major I, and Boötes I), which are near enough to allow deep observations but subtend a large angular area and would require a large number of pointings with JWST or HST.

6 SUMMARY

Using simple Monte Carlo simulations, we have explored the statistical challenges of constraining the sub-stellar IMF in Local Group dwarf galaxies with direct star counts. We consider both the idealized case in which stellar masses are known exactly and the IMF can be fit directly, and the more realistic case in which the IMF is determined by comparing model CMDs to photometry (see Appendix C). Our main results are as follows.

- (i) *Statistical constraints on the turnover in the IMF:* In order to break the degeneracy between the characteristic mass (m_c) and width (σ) when fitting a lognormal IMF, observations must reach a minimum stellar mass $M_{\text{obs}} \sim m_c$; i.e., approximately $M_{\text{obs}} = 0.2M_\odot$ if the IMF is similar to the Chabrier IMF measured in the MW (see Figure 2). Stronger constraints can be obtained for shallower observations, but this typically requires a prohibitively large sample of stars. Existing HST observations have not yet reached sufficiently low masses to jointly constrain m_c and σ .
- (ii) *Fitting a power law to masses drawn from a lognormal IMF yields shallow power law slopes:* The degeneracies of fitting a lognormal IMF can be avoided by fitting a single power law $\xi(m) \propto m^{-\alpha}$. However, if observations can only sample the IMF in a narrow mass range such as $0.5M_\odot \lesssim m \lesssim 0.8M_\odot$, which is typical for attempts to constrain the IMFs of MW satellites with HST, this will yield a power law slope $\alpha \sim 1.55$, even when masses are drawn from a Chabrier IMF (see Figure 3). This slope is different from the canonical slope measured in the MW ($\alpha = 2.3$ for a Kroupa IMF, or $\alpha = 2.35$ for a Salpeter IMF), simply because of the mass range probed by observations. Existing reports of systematically shallower IMFs in MW satellite galaxies may thus in fact be consistent with a Chabrier IMF.
- (iii) *Fixing IMF parameters during fitting can lead to underestimated uncertainty:* A common approximation in fitting a lognormal IMF is to fix the parameter σ to e.g. the Chabrier value, in order to avoid the degeneracy with m_c . However, this practice should be avoided, as it will dramatically underestimate the uncertainty in m_c (see Figure 5). This is not necessarily a problem if the obtained constraints on m_c are only compared to other constraints obtained in the same

mass range with the same fixed σ value, but it can lead to erroneous indications of IMF variations if values of m_c are compared across different studies.

- (iv) *Observations required to detect variability in the low-mass IMF:* If the slope of the subsolar IMFs of different galaxies are similar at $m > 0.5M_\odot$ (as appears to be the case), observations must reach down to at least m_c in order to detect variability in the IMF and rule out the possibility that observed data are drawn from a Chabrier IMF (see Figure 6). In addition, failure to detect a turnover in the IMF does not imply deviations from the Chabrier IMF unless observations reach down to $m \sim 0.2M_\odot$ (see Figure 4).
- (v) *Prospects for constraining the IMFs of Local Group dwarf galaxies with JWST:* Near-infrared observations with JWST will probe significantly lower stellar masses at fixed distance than previous studies with HST (see Figure 7), which will allow for unprecedented measurement of the low mass turnover (if it exists) in the IMFs of nearby dwarf galaxies. JWST observations will be complete down to the hydrogen burning limit for galaxies within ~ 60 kpc, and down to $\sim 0.22M_\odot$ for galaxies at distances of ~ 150 kpc. Promising targets for IMF studies in the Local Group are highlighted in Figure 8. JWST observations of most nearby dwarf galaxies will not be significantly limited by crowding, but a similar number of pointings will be required per galaxy as with HST observations to build a sufficiently large sample of stars (see Figure 9).

ACKNOWLEDGEMENTS

We thank the anonymous referee for a constructive report which improved this paper. We thank Marla Geha and Clement Wagner for comments on the manuscript, and Hans-Walter Rix and Kevin Covey for helpful discussions. KE gratefully acknowledges support from a Berkeley graduate fellowship, a Hellman award for graduate study, and an NSF Graduate Research Fellowship. EQ was supported by NASA ATP grant 12-ATP-120183, a Simons Investigator award from the Simons Foundation, and the David and Lucile Packard Foundation. We ran numerical calculations on the Caltech compute cluster “Zwicky” (NSF MRI award #PHY-0960291). This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013), and Matplotlib (Hunter 2007).

REFERENCES

- Allard F., Hauschildt P. H., Alexander D. R., Starrfield S., 1997, *ARA&A*, **35**, 137
- Astropy Collaboration et al., 2013, *A&A*, **558**, A33
- Bastian N., Covey K. R., Meyer M. R., 2010, *ARA&A*, **48**, 339
- Bessell M. S., Stringfellow G. S., 1993, *ARA&A*, **31**, 433
- Bochanski J. J., Hawley S. L., Covey K. R., West A. A., Reid I. N., Golimowski D. A., Ivezić Ž., 2010, *AJ*, **139**, 2679
- Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, **427**, 127
- Brown T. M., et al., 2012, *ApJ*, **753**, L21
- Cappellari M., et al., 2012, *Nature*, **484**, 485
- Chabrier G., 2003a, *PASP*, **115**, 763
- Chabrier G., 2003b, *ApJ*, **586**, L133
- Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *ApJ*, **823**, 102
- Conroy C., van Dokkum P. G., 2012, *ApJ*, **760**, 71
- Coulter D. A., et al., 2017, *ApJ*, **835**, 183
- Covey K. R., et al., 2008, *AJ*, **136**, 1778
- Dalcanton J. J., et al., 2012, *ApJS*, **200**, 18
- Dolphin A. E., 2002, *MNRAS*, **332**, 91
- Dotter A., 2016, *ApJS*, **222**, 8
- Elmegreen B. G., 1999, *ApJ*, **515**, 323
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Frebel A., Norris J. E., 2015, *ARA&A*, **53**, 631
- Frebel A., Norris J. E., Gilmore G., Wyse R. F. G., 2016, *ApJ*, **826**, 110
- Geha M., et al., 2013, *ApJ*, **771**, 29
- Gelman A., Rubin D., 1991, in *Bayesian Statistics 4*. pp 625–631
- Gelman A., Carlin J., Stern H., Dunson D., Vehtari A., Rubin D., 2013, *Bayesian Data Analysis: Third Edition*. CRC Press
- Grillmair C. J., et al., 1998, *AJ*, **115**, 144
- Grossman A. S., 1970, *ApJ*, **161**, 619
- Hartwick F. D. A., 1976, *ApJ*, **209**, 418
- Hernandez X., 2012, *MNRAS*, **420**, 1183
- Hopkins P. F., 2013, *MNRAS*, **433**, 170
- Hunter J. D., 2007, *Computing In Science & Engineering*, **9**, 90
- Ivezić Ž., Connolly A. J., VanderPlas J. T., Gray A., 2014, *Statistics, Data Mining, and Machine Learning in Astronomy*
- Kalirai J. S., et al., 2013, *ApJ*, **763**, 110
- Kirby E. N., Simon J. D., Geha M., Guhathakurta P., Frebel A., 2008, *ApJ*, **685**, L43
- Kirby E. N., Lanfranchi G. A., Simon J. D., Cohen J. G., Guhathakurta P., 2011, *ApJ*, **727**, 78
- Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, *ApJ*, **779**, 102
- Kroupa P., 2001, *MNRAS*, **322**, 231
- Kroupa P., 2002, *Science*, **295**, 82
- Kroupa P., Gilmore G., Tout C. A., 1991, *MNRAS*, **251**, 293
- Kroupa P., Weidner C., Pflamm-Altenburg J., Thies I., Dabringhausen J., Marks M., Maschberger T., 2013, *The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations*. p. 115, doi:10.1007/978-94-007-5612-0_4
- Krumholz M. R., 2011, *ApJ*, **743**, 110
- Krumholz M. R., Myers A. T., Klein R. I., McKee C. F., 2016, *MNRAS*, **460**, 3272
- Kumar S. S., 1963, *ApJ*, **137**, 1121
- Lai D. K., Lee Y. S., Bolte M., Lucatello S., Beers T. C., Johnson J. A., Sivarani T., Rockosi C. M., 2011, *ApJ*, **738**, 51
- Larson R. B., 1998, *MNRAS*, **301**, 569
- Liddle A. R., 2007, *MNRAS*, **377**, L74
- Liebert J., Probst R. G., 1987, *ARA&A*, **25**, 473
- Maíz Apellániz J., Úbeda L., 2005, *ApJ*, **629**, 873
- Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, *A&A*, **482**, 883
- Maschberger T., Kroupa P., 2009, *MNRAS*, **395**, 931
- McConnachie A. W., 2012, *AJ*, **144**, 4
- McConnell N. J., Lu J. R., Mann A. W., 2016, *ApJ*, **821**, 39
- Miller G. E., Scalo J. M., 1979, *ApJS*, **41**, 513
- Narayanan D., Davé R., 2012, *MNRAS*, **423**, 3601
- Offner S. S. R., Clark P. C., Hennebelle P., Bastian N., Bate M. R., Hopkins P. F., Moraux E., Whitworth A. P., 2014, *Protostars and Planets VI*, pp 53–75
- Olver S., Townsend A., 2013, preprint, (arXiv:1307.1223)
- Padoan P., Nordlund A., Jones B. J. T., 1997, *MNRAS*, **288**, 145
- Pagel B. E. J., 1997, *Nucleosynthesis and Chemical Evolution of Galaxies*

- Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, *ApJS*, **192**, 3
- Paxton B., et al., 2013, *ApJS*, **208**, 4
- Paxton B., et al., 2015, *ApJS*, **220**, 15
- Ross T. L., Holtzman J., Saha A., Anthony-Twarog B. J., 2015, *AJ*, **149**, 198
- Salpeter E. E., 1955, *ApJ*, **121**, 161
- Scalo J. M., 1986, *Fundamentals Cosmic Phys.*, **11**, 1
- Schwarz G., 1978, *Ann. Statist.*, **6**, 461
- Simon J. D., Geha M., 2007, *ApJ*, **670**, 313
- Smith R. J., 2014, *MNRAS*, **443**, L69
- Spergel D., et al., 2013, preprint, ([arXiv:1305.5422](https://arxiv.org/abs/1305.5422))
- Tolstoy E., Hill V., Tosi M., 2009, *ARA&A*, **47**, 371
- Vargas L. C., Geha M., Kirby E. N., Simon J. D., 2013, *ApJ*, **767**, 134
- Webster D., Bland-Hawthorn J., Sutherland R., 2015, *ApJ*, **799**, L21
- Weisz D. R., et al., 2013, *ApJ*, **762**, 123
- Weisz D. R., Dolphin A. E., Skillman E. D., Holtzman J., Gilbert K. M., Dalcanton J. J., Williams B. F., 2014, *ApJ*, **789**, 147
- Wolf J., Martinez G. D., Bullock J. S., Kaplinghat M., Geha M., Muñoz R. R., Simon J. D., Avedo F. F., 2010, *MNRAS*, **406**, 1220
- Wyse R. F. G., Gilmore G., Houdashelt M. L., Feltzing S., Hebb L., Gallagher III J. S., Smecker-Hane T. A., 2002, *New Astron.*, **7**, 395
- van Dokkum P. G., 2008, *ApJ*, **674**, 29
- van Dokkum P., Conroy C., Villaume A., Brodie J., Romanowsky A., 2016, preprint, ([arXiv:1611.09859](https://arxiv.org/abs/1611.09859))

APPENDIX A: DRAWING MASSES FROM THE IMF

Masses are drawn from the IMF using inverse transform sampling (e.g. [Olver & Townsend 2013](#)). Let $\xi(m)$ be an arbitrary IMF, normalized such that $\int_{M_{\min}}^{M_{\max}} \xi(m) dm = 1$. We define the cumulative initial mass function $N(m) = \int_{M_{\min}}^m \xi(m') dm'$. Then $N(m)$ gives the probability that a random draw from the IMF will be less than m . By definition, if m_i are masses drawn from $\xi(m)$, then $N(m_i)$ will be uniformly distributed; that is, $N(m_i) = \mathcal{U}(0, 1)$.

We define the inverse cumulative mass function $N^{-1}(x)$, which takes a number $x \in [0, 1]$ and returns the mass m for which the probability of drawing a mass less than m from $\xi(m)$ is x . Thus, if u_i are random draws from a uniform distribution $\mathcal{U}(0, 1)$, then $N^{-1}(u_i) = N^{-1}(N(m_i)) = m_i$ will be random draws from $\xi(m)$. Therefore, to draw masses from $\xi(m)$, one can draw values from $\mathcal{U}(0, 1)$ and pass them through $N^{-1}(x)$.

For a lognormal IMF in the form of Equation 1, the inverse of the cumulative distribution is given by

$$N^{-1}(x) = 10^{\left(\sqrt{2}\sigma \text{erf}^{-1}[g(M_{\max})x + (1-x)g(M_{\min})] + \log m_c\right)}, \quad (\text{A1})$$

where $g(y) = \text{erf}\left[(\log y - \log m_c)/(\sqrt{2}\sigma)\right]$ and $\text{erf}^{-1}(x)$ is the inverse error function.

APPENDIX B: TRANSLATING BETWEEN MASSES AND MAGNITUDES

We convert between stellar masses and absolute magnitudes by interpolating on a grid of stellar models. For our primary

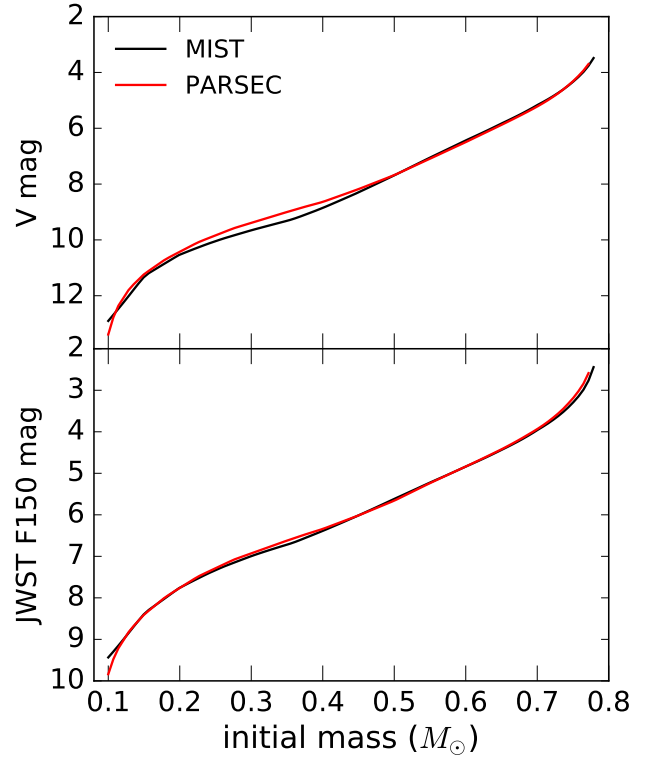


Figure B1. Absolute magnitude in the *V* (top) and JWST *F150* (bottom) bands, as predicted by two different stellar evolution codes. Both models are for stars of age 12.5 Gyr and metallicity $[\text{Fe}/\text{H}] = -2$. We use the MIST models for our primary analysis.

analysis, we assume a single stellar population with age 12.5 Gyr and metallicity $[\text{Fe}/\text{H}] = -2$. For a particular stellar evolutionary model, this uniquely specifies the transformation between stellar initial mass and absolute magnitude. We do not attempt to account for complications due to α -element enhancement, stellar rotation, or binary evolution, which for our purposes represent small second order corrections.

Figure B1 shows the relation between initial mass and absolute magnitude predicted by stellar evolution models. We compare the relations predicted by two different codes: MIST ([Choi et al. 2016](#); [Dotter 2016](#)), which is based on the MESA code ([Paxton et al. 2011, 2013, 2015](#)), and PARSEC ([Bressan et al. 2012](#)), which is based on the Padova isochrones ([Marigo et al. 2008](#)). We plot mass-luminosity relations in both the *V* (top) and the JWST *F150W* (bottom) bands.

The two models are, for the most part, in good agreement. The most significant deviation occurs in the *V* band for stars with initial masses in the range $(0.3-0.4)M_{\odot}$, where the two models disagree by up to 10% on the initial mass corresponding to a given absolute magnitude. This disagreement is likely a result of the sensitivity of the cool atmospheres of low-mass stars to molecular opacity ([Allard et al. 1997](#)). The two models are in better agreement at the same mass in the redder *F150W* band, which is less strongly affected by molecular absorption in stellar atmospheres.

We emphasize that our main results are all derived by fitting masses directly, without transforming into magnitude

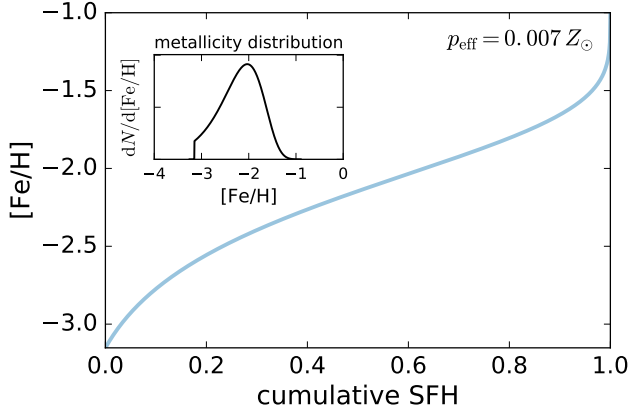


Figure C1. Pre-enriched “leaky box” chemical evolution model assumed in creating the CMD in Figure C2. This model predicts a metallicity distribution function consistent with observations of many nearby low-mass galaxies.

space. Uncertainties in stellar models thus do not directly affect our calculations of the IMF constraints that can be obtained for a given M_{obs} and M_{star} ; they only come into play when these values are translated into magnitudes. On the other hand, if fitting is done directly in magnitude space as in the next section, errors in stellar models could introduce biases in the recovered IMF.

APPENDIX C: EXTRACTING THE IMF FROM A CMD

In this section, we summarize how the IMF can be recovered from a CMD if the masses of individual stars are not known exactly. We first generate a “data” CMD by assigning magnitudes to masses sampled from the IMF and applying a completeness function. We then generate “model” CMDs by forward modeling IMFs with different parameters, choosing the best-fit IMF as the one for which the model CMD most closely matches the data CMD. Our goal is primarily to demonstrate that this approach, which is commonly used in observational studies, does not introduce systematic errors compared to our fiducial method of fitting masses directly, which was described in Section 2.3.

We assume stellar magnitudes have been measured in two filters, which we label V and I . Of course, this analysis is in principle applicable to CMDs generated in any two filters. We calculate V and I magnitudes for stars of a given initial mass, metallicity, and age by interpolating on a grid of MIST stellar models. In order to minimize the weight given to stars in poorly understood phases of stellar evolution (which are also not sensitive to the IMF), we include only main-sequence stars in our fit.

C1 Stellar metallicity distribution function

Since stars of a given initial mass will have different magnitudes and colors at different metallicities, it is important to realistically model the metallicity distribution function (MDF) when generating a CMD. Nearby ultra-faint and

dwarf spheroidal galaxies exhibit a significant range of stellar metallicities, even in galaxies with simple star formation histories dominated by old (> 12 Gyr) stars. Observations of dwarf galaxies in the Local Group (e.g. Simon & Geha 2007; Kirby et al. 2008, 2011; Lai et al. 2011; Kirby et al. 2013; Ross et al. 2015; Frebel et al. 2016) generally find a mean metallicity of $[\text{Fe}/\text{H}] \approx -2$, with a range of metallicities as large as ~ 2 dex (i.e., $-3 \lesssim [\text{Fe}/\text{H}] \lesssim -1$ for stars in a single galaxy).

We use the MDF predicted by a “leaky box” analytic chemical evolution model (Hartwick 1976; Pagel 1997) to generate our CMDs. We modify the model slightly to allow for pre-enrichment, so that the most metal-poor stars have metallicity Z_0 (where we set $Z_0 = 10^{-3} Z_{\odot}$) rather than zero metallicity. The MDF predicted by this model has been shown to be in good agreement with the observed MDFs of many Local Group dwarf spheroidal and ultra-faint galaxies (Kirby et al. 2011, 2013). It is parameterized by a mass-loading factor η , which represents the gas mass lost to outflows per unit mass of stars formed, and the nucleosynthetic yield p , which represents the mass of metals returned to the ISM per unit mass of stars formed. p depends only on the IMF and the details of nuclear burning, whereas η depends on a galaxy’s potential depth and star formation history.

If the highest metallicity star formed by $z = 0$ has metallicity $Z = Z_1$, the fraction of stars with metallicity less than Z is given by

$$\frac{M_{\star}(< Z)}{M_{\star}(< Z_1)} = \frac{1 - \exp[(Z_0 - Z)/p_{\text{eff}}]}{1 - \exp[(Z_0 - Z_1)/p_{\text{eff}}]}, \quad (\text{C1})$$

where $p_{\text{eff}} = p/(1+\eta)$ is the effective yield. The corresponding MDF is

$$\frac{dN}{d[\text{Fe}/\text{H}]} \propto 10^{[\text{Fe}/\text{H}]} \exp \left[Z_{\odot} \frac{10^{[\text{Fe}/\text{H}]_0} - 10^{[\text{Fe}/\text{H}]} }{p_{\text{eff}}} \right], \quad (\text{C2})$$

in the region $Z_0 \leq Z \leq Z_1$, and 0 outside this region. Here we have defined $[\text{Fe}/\text{H}]_0 = \log(Z_0/Z_{\odot})$.

Figure C1 shows the enrichment history and metallicity distribution predicted for $p = 0.01$ and $\eta = 100$. These values were chosen to produce a mean metallicity $\bar{Z} = 0.01 Z_{\odot}$, comparable to what is found in observed galaxies. We use these values and the MDF shown in Figure C1 to generate our CMDs.

C2 Completeness

Thus far, we have assumed that observations can resolve all stars with masses above M_{obs} and none of the stars with masses below M_{obs} . In real observations, this limit is manifest as a gradual drop-off in completeness with increasing magnitude, not as a sharp cutoff at a particular stellar mass.

We model this drop-off with a completeness function $c(M_V)$, which represents the fraction of stars with magnitude M_V (which can, in principle, be measured in any filter) that can be observed. We use the completeness function

$$c(M_V) = \frac{1}{2} \left[1 - \tanh(M_V - M_{V,0.5}) \right]. \quad (\text{C3})$$

Here, $M_{V,0.5}$ is the magnitude at which completeness falls to 50%, which is chosen for a given M_{obs} by interpolating on a grid of stellar models. We use this function because it provides a reasonably good approximation for the

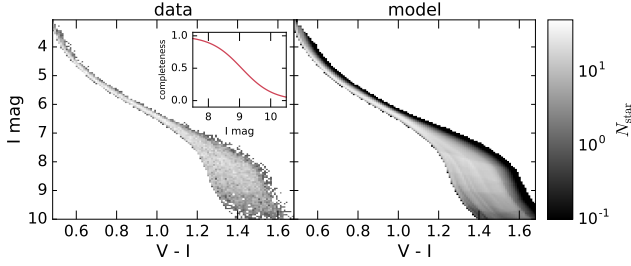


Figure C2. *Left*: synthetic CMD generated for a stellar population with $M_{\text{obs}} = 0.2 M_{\odot}$ and $M_{\text{star}} = 10^4 M_{\odot}$. Only main sequence stars are shown. *Right*: best-fit model CMD obtained by maximizing the likelihood in Equation C4. Both the model and the synthetic CMD assume a stellar population of age 12.5 Gyr with a realistic pre-enriched leaky box metallicity distribution function. Inset shows the adopted I -band completeness function.

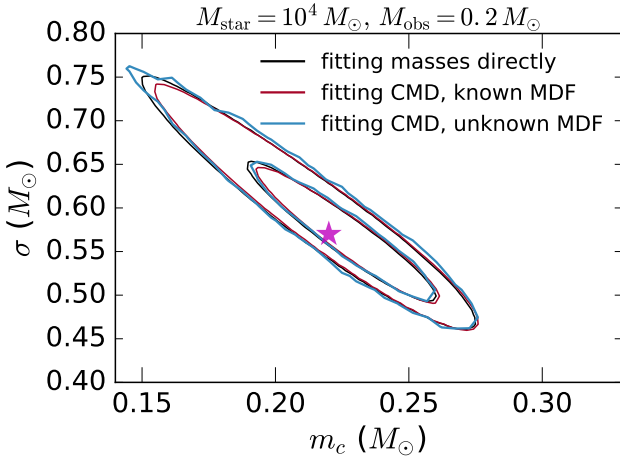


Figure C3. Comparison of IMF fits obtained for an example stellar population. Contours show 1 and 2σ constraints on the parameters of the IMF obtained with different fitting methods. **Black**: we fit masses drawn from the IMF directly, as in Equation 2. **Red**: We generate a synthetic CMD with a known leaky-box metallicity distribution function (MDF) and fit using Equation C4. **Blue**: We again fit the synthetic CMD, but now the parameters of the MDF are unknown; contours are generated by marginalizing over the MDF parameters. Star shows the true parameters of the IMF from which masses are drawn.

empirically-estimated completeness functions found in observational IMF studies of low-mass galaxies (Wyse et al. 2002; Kalirai et al. 2013). The function for $M_{\text{obs}} = 0.2 M_{\odot}$ is shown in the inset of Figure C2.

C3 Fitting the CMD

We generate a “data” CMD by sampling masses from the IMF and metallicities from the MDF, transforming into magnitude space using the MIST stellar models, and discarding stars of magnitude M_V with probability $1 - c(M_V)$.

Following the method outlined in Dolphin (2002), we then divide the CMD into N bins in both I magnitude and $V - I$ color, so that the full pixelated CMD (or “Hess dia-

gram”) contains N^2 total pixels. We choose $N = \sqrt{N_{\text{star}}}$, so that the number of pixels is equal to the number of stars. We have verified that our fit is not sensitive to the choice of bins, as long as bins are small enough to capture the smallest CMD features.

We generate model CMDs for a population with a given IMF by linearly combining isochrones of different metallicities, weighting points along each isochrone by the IMF and weighting different isochrones by the MDF. We calculate the likelihood for a given model by summing over all pixels, where the likelihood in a given pixel is set by Poisson statistics. Explicitly, if m_i represents the i -th pixel of a given model CMD and d_i represents the same pixel in the data CMD, the log-likelihood can be written as

$$\ln \mathcal{L} = \sum_{m_i \neq 0} [d_i \ln m_i - m_i - \ln(d_i!)] . \quad (\text{C4})$$

The sum is over all pixels in which the model probability is non-zero. For pixels with $d_i > 100$, where it is computationally impractical to evaluate the factorial exactly, we use Stirling’s approximation to evaluate $\ln(d_i!)$. For $d_i = 100$, this produces a negligible error of a few $\times 10^{-4}$ percent; the fractional error declines further with increasing d_i . We note that while d_i must be a non-negative integer, m_i need not be an integer.

In Figure C2, we compare a synthetic “data” CMD generated for masses drawn from a Chabrier IMF in the mass range $m = (0.2 - 0.77) M_{\odot}$ to the best-fit model CMD generated by maximizing the likelihood in Equation C4. The model essentially recovers the “data” CMD exactly: the apparent differences in the regions where pixels in the model are dark are due to the fact that the model predicts significantly less than 1 star per bin, while the same pixels in the data CMD must have an integer number of stars (typically 0 or 1).

Figure C3 shows the probability contours for m_c and σ generated by this fitting procedure for a sample of masses drawn from a Chabrier IMF in the mass range $m = (0.2 - 0.77) M_{\odot}$. V and I -band magnitudes were generated assuming a uniform age of 12.5 Gyr and a pre-enriched leaky box MDF. We considered two cases: one in which the parameters of the MDF are known a priori, and one in which we simultaneously fit for the parameters of the MDF and the IMF. In the latter case, the plotted IMF probability contours represent the results of marginalizing over the MDF parameters.

The constraints on m_c and σ obtained by fitting the synthetic CMD are in excellent agreement with those obtained by fitting the IMF directly to the masses as described in Section 2.3. This remains true even when the parameters of the MDF are not known a priori and the MDF is fit simultaneously with the IMF. In other words, there is no loss of information when fitting in color-magnitude space compared to fitting masses directly, even though initial mass does not uniquely determine the magnitude of an individual star for a population with a spread of metallicities. This is because the CMD contains information about the distribution of stellar metallicities: the color distribution of the main sequence is a direct transformation of the MDF.

This suggests that fitting a CMD provides equally strong constraints on the IMF as fitting masses directly — in the limit of perfect observations. It is important to re-

member, however, that this exercise is still highly idealized. In practice, the precision of IMF constraints from CMD fitting could also be limited by photometric uncertainties (see [Weisz et al. 2013](#)), by uncertainties in the stellar models used to translate between mass and magnitude, and by degeneracies between the binary fraction and the MDF. As we have not attempted to model these uncertainties, our IMF constraints represent best-case upper limits.